Journal of the American Association for Laboratory Animal Science
Copyright 2012
by the American Association for Laboratory Animal Science

Vol 51, No 1
January 2012
Pages 31–6

Validation of a Body Condition Scoring System in Rhesus Macaques (*Macaca mulatta*): Inter- and Intrarater Variability

Karen J Clingerman[1,*] and Laura Summers[2]

Body condition scoring (BCS) is a subjective semiquantitative method of assessing body fat and muscle. Scoring systems use a scale in which the midrange represents optimal body condition, lower values represent lean to emaciated conditions, and higher values indicate excessive body fat. A valid BCS system is clearly described, relevant to the species, shows agreement within and between raters, and is consistent with objective measures. The goal of the current study was to assess intra- and interrater variability of a BCS system that uses a 1-to-5 scale and entails the palpation of key anatomic sites (hips, spine, pelvis, thorax, and abdomen) to assess prominence of bony structures, muscle mass, and subcutaneous fat. To assess interrater variability, 4 raters independently assessed BCS in 616 rhesus macaques (*Macaca mulatta*) in 4 age groups: infant, younger than 1 y; juvenile, 1 to 4 y; subadult, 4 to 7 y; and adult, 7 to 17 y. To assess intrarater variability, each rater independently reevaluated a subset of adult macaques (n = 15) within 2 wk of initial evaluation. A weighted κ score was used to analyze intra- and interrater variability. Agreement between raters was highest for subadult and adult macaques, intermediate for juveniles, and least for infants. Intrarater agreement was high for all raters except one, for which it was moderate. Our results suggest that raters applied the BCS system most consistently to adult and subadult macaques and less so to juvenile and infant animals. However, the percentage agreement between raters to within one half of a score unit increased markedly when raters scored infants in the context of 'as is' rather than 'ideal for age.'

Abbreviation: BCS, body condition score.

Body condition scoring (BCS) is a subjective, semiquantitative method of assessing body fat and muscle.[2] Scoring of body condition is performed in a wide variety of species including sheep, cattle, horses, dogs, cats, rats, and mice.[2,6,8,11,16-19,22,25-27,30] BCS can be used to assess overall health, production, and dietary management and can be a predictive factor in disease risk and outcome.[7-12,14,21,22-24,29] Scales for scoring systems typically range from 1 to 5, 1 to 6, or 1 to 9, with midrange scores representing optimal body condition, lower values representing lean or emaciated conditions, and higher values indicating excessive body fat. With practice, most protocols are easy to learn and apply, especially when the scale is well described. Scoring of animals can be incorporated readily into physical examination procedures and can be useful in assessing the health and nutrition of individual animals.

Although a scoring system that is based on visual assessment of free-ranging rhesus monkeys has been described,[1] the system used in the current study is similar to those used in other species which are based on a hands-on assessment of the animals. The system (Figure 1), which has been described in detail elsewhere,[3] uses a scale of 1 to 5 in half units. In brief, the system entails the palpation of key anatomical sites (hips, spine, pelvis, thorax, and abdomen) to assess prominence of bony structures, muscle mass, and subcutaneous fat.

Valid BCS systems have several key characteristics. First, the scale is well described and relevant to the species to which it is applied; these features have already been addressed for the system we use here.[3] In addition, users who apply the scoring system accurately independently score the same animal similarly. Furthermore, provided that the animal's body condition hasn't changed, a rater assigns the same score whenever the same animal is presented. Finally, valid BCS systems are consistent with other objective methods of assessing body condition. The goals of the current study were to assess interrater and intrarater variability. The questions posed were first, can the agreement between independent raters be explained by more than just chance, and second, can the scoring system be applied consistently by each rater.

Materials and Methods

Animals. All macaques were housed in large outdoor enclosures at the California National Primate Research Center (Davis, CA). Each 1.5-acre enclosure housed 50 to 175 rhesus macaques (*Macaca mulatta*; age, newborn to 24 y). Macaques were provided water and a commercial diet (Purina Hi-Pro Monkey Chow, PMI, St Louis, MO) ad libitum and supplemented with fresh fruits and vegetables 2 times each week. Macaques were maintained in 1 of 2 types of colonies—conventionally reared and SPF—within the outdoor enclosures. Conventionally reared animals were negative for SIV, simian retrovirus, and simian transmissible lymphoma virus. Level 1 SPF macaques were negative for the agents previously listed and *Macacine herpesvirus* 1; level 3 SPF animals were also negative for cytomegalovirus, foamy virus, and rhesus rhabdovirus. SPF macaques in outdoor enclosures were screened annually. All procedures were approved by the IACUC, facilities were AAALAC-accredited, and all procedures

Received: 14 Mar 2011. Revision requested: 13 Apr 2011. Accepted: 12 Aug 2011.
[1]Department of Animal Resources, The Scripps Research Institute, La Jolla and [2]California National Primate Research Center, University of California, Davis, California.
[*]Corresponding author. Email: karenc@scripps.edu

1	**EMACIATED** – Very prominent hip bones (easily palpable and likely visible), prominent facial bones, spinous processes and ribs. Minimal to no muscle mass is palpable over ileum or ischium. Anus may be recessed between ischial callosities. Body is very angular, no subcutaneous fat layer to smooth out prominences.
1.5	**VERY THIN** – Hips, spinous processes, and ribs are prominent. Facial bones may be prominent. There is very little muscle present over the hips and back. Anus may be recessed between ischial callosities. Body is angular, no subcutaneous fat to smooth out prominences
2	**THIN** – Very minimal fat reserves, prominent hip bones and spinous processes. Hips, spinous processes and ribs are easily palpable with only a small amount of muscle mass over hips and lumbar region.
2.5	**LEAN** – Overlying muscle gives hips and spine a more firm feel. Hip bones and spinous processes are readily palpable, but not prominent. Body is less angular because there is a thin layer of subcutaneous fat.
3	**OPTIMUM** –Hip bones, ribs and spinous processes are palpable with gentle pressure but generally not visible. Well developed muscle mass and subcutaneous fat layer gives spine and hips smooth but firm feel. No abdominal, axillary or inguinal fat pads.
3.5	**SLIGHTLY OVERWEIGHT** – Hip bones and spinous processes palpable with firm pressure but are not visible. Bony prominences smooth. Rib contours are smooth and only palpable with firm pressure. Small abdominal fat pad may be present.
4	**HEAVY** – Bony contours are smooth and less well defined. Hip bones, spinous processes and ribs may be difficult to palpate due to more abundant subcutaneous fat layer. May have fat deposits starting to accumulate in the axillary, inguinal or abdominal areas.
4.5	**OBESE** – This animal will often have prominent fat pads in the inguinal, axillary or abdominal region. Abdomen will be pendulous when animal sitting or ambulating. Hip bones and spinous processes difficult to palpate. Bony contours smooth and poorly defined.
5	**GROSSLY OBESE** –Obvious, large fat deposits in the abdominal, inguinal and axillary regions. Abdominal palpation is very difficult due to large amount of mesenteric fat. Pronounced fat deposits may alter posture/ambulation. Hip bones, rib contours and spinous processes only palpable with deep palpation.

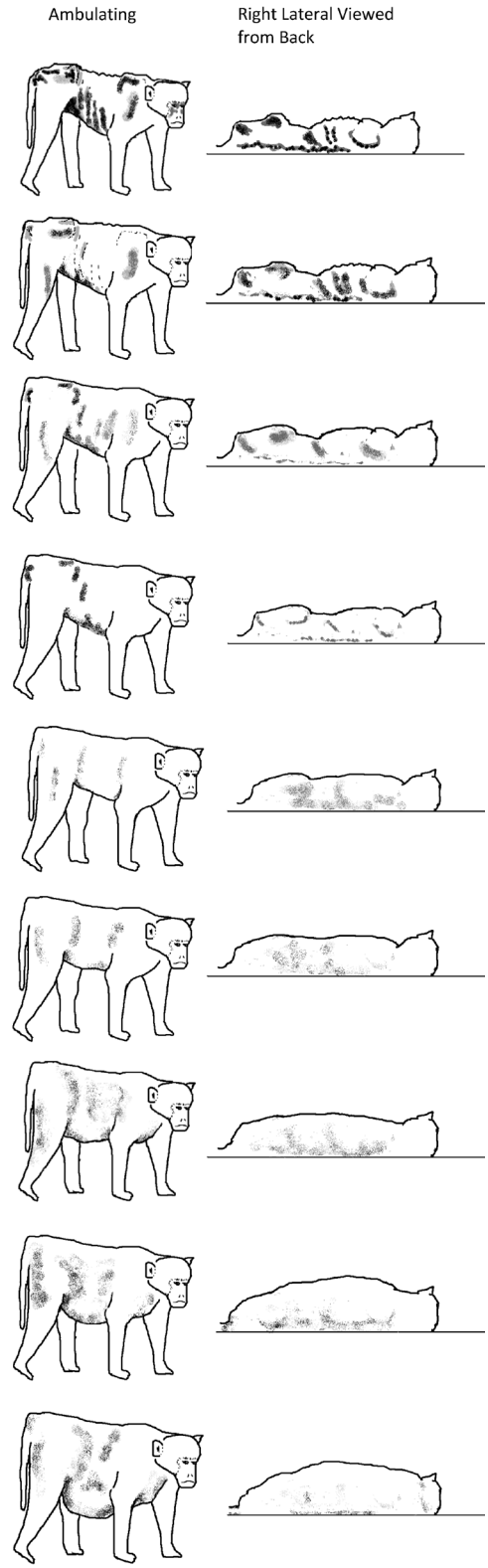

Figure 1. Body condition scoring of nonhuman primates, using *Macaca mulatta* as a model. Image reproduced from reference 3 with permission.

involving animals adhered to recommendations in the *Guide for the Care and Use of Laboratory Animals*.[13]

Raters. Four veterinarians, both staff and resident, from 2 institutions were assigned as raters. Individual training consisted of becoming familiar with the scoring system[3] and using it during routine physical examinations that they performed. We elected to not standardize or formalize training, because we wanted to determine whether after reading the description and applying the scoring system, raters would interpret it similarly.

Body condition scoring. ***Assessing interrater variability.*** Four raters independently evaluated all animals and assigned a score

to each. As they were immobilized and presented to the raters, macaques were evaluated and scored based on the previously described scoring system[3]. In brief, each animal was assessed by palpation of bony prominences, muscle, and fat deposition over key body areas including the thorax, spine, hips, and abdomen. Scores ranged from 1, indicating an emaciated animal, to 5, which represents a grossly obese animal (Figure 1), in half-unit intervals.

Assessing intrarater variability. The same 4 raters independently reevaluated a subset of adult animals by using the same scoring criteria described. The colony manager randomly selected these animals from 2 different outdoor enclosures; raters were blind to the animals' previous histories, including BCS. To avoid any potential significant changes in body condition over time, macaques were rescored within 2 wk of the initial evaluation.

Treatment groups. Four corrals of animals were evaluated (n = 616). All body condition evaluations were done in conjunction with the regularly scheduled immobilization and physical examination of all animals within the outdoor enclosure. Animals were immobilized in no particular order by CNPRC personnel with ketamine (Fort Dodge Animal Health, Fort Dodge, IA) administered intramuscularly within a standard dose range of 10 to 15 mg/kg. Animals to be rescored by raters were randomly selected from 2 outdoor enclosures (n = 15) and transferred temporarily to indoor housing. The animals were immobilized with ketamine administered intramuscularly at a dose of 10 mg/kg.

Data analysis. The animal data was divided into relevant age groupings of infant (younger than 1 y), juvenile (1 to 4 y), subadult (4 to 7 y), and adult (7 to 17 y). In some cases, the actual ages of macaques differed slightly from those of the category assigned. Natural breaks in the age groupings of the macaques and those whose birthdates put them closer to the next age category accounted for the differences.

A weighted κ statistic with quadratic weights was used to analyze intrarater and interrater variability. The weighted κ gives greater weight to those scores that are closer to each other than to more discrepant scores. A significant κ score ($P < 0.05$) rejects the null hypothesis of no agreement between raters. In addition, the κ score can be interpreted based on level of agreement. The following standards have been proposed for strength of agreement according to the κ coefficient and were adopted for the current study: 0 or less, poor agreement; 0.10 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect.[20] Interrater variability was analyzed over all raters for each age group. Intrarater variability was analyzed over 15 macaques randomly selected from the 7- to 17-y age group that were rescored within 1 to 2 wk of the initial scoring.

Results

Overall, more female macaques than male were assessed (Table 1). The composition of the groups housed in the large outdoor enclosures parallels the social structure of wild rhesus populations and, as such, the sex ratio is skewed in favor of female macaques.

The frequency distribution of each score within each age group is presented in Table 2.The most frequently assessed score (mode) was 2.0, 2.5, 3.0, and 3.5 for infants, juveniles, subadults, and adults, respectively. The ranges of scores reported for each age category was 1 to 3 for infants, 1.5 to 3.5 for juveniles, 1.5 to 4.5 for subadults, and 1 to 5 for adults with all scores in the range represented.

The percentage agreement between raters was assessed for both complete agreement between all raters (that is, all raters assigned the same score to a particular animal) and for agreement to within 0.5 score unit. Agreement to within 0.5 unit occurred in 1 of 2 ways: either 3 raters assigned the same score and 1 assigned a score that was within 0.5 unit of the score of the other 3 raters, or 2 raters assigned the same score and the 2 remaining raters assigned a different score that is the same for both of them but differed from that of the first 2 raters by 0.5. In each of these cases, scores across all raters were always within 0.5. Rater agreement to within 0.5 was 23% for the infant group, 63% for juveniles, 83% for the subadult group, and 84% for adults (Figure 2).

Overall agreement between scorers was highest for subadult and adult macaques (Table 3). In pairwise comparisons between raters, the κ values were largely concordant across the raters. In comparisons across age groups, κ values increased as the age of the macaques increased, consistent with overall percentage agreement. Intrarater agreement to within 0.5 score unit was 100% for rater 1 and 93.3% for raters 2, 3, and 4 (Figure 3). The κ values (SE) were significant ($P < 0.05$) for each intrarater comparison (0.774 [0.249], 0.583 [0.237], 0.636 [0.224], and 0.791 [0.252]) for raters 1, 2, 3, and 4, respectively. According to standard definitions regarding levels of agreement,[20] interrater agreement was substantial for subadult and adult animals, fair for juveniles, and slight for infants. Intrarater agreement was substantial for all raters but one, for which it was moderate.

The frequency distributions of pairwise comparisons between raters are presented in Table 4 For each age group, the majority of the scores fell within 0.5 score unit. Scoring discrepancies were more numerous for infants and juveniles compared with adults and subadults. In addition, several scoring tendencies can be inferred from Table 4. Overall, rater 1 tended to score animals higher across all age groups. Raters 1 and 2 tended to score juvenile animals higher than did raters 3 and 4. For infants, rater 1 tended to score animals higher than did all other raters. Agreement between raters was similar across pairs for adults and subadults. For juveniles, raters 1 and 2 tended to agree and raters 3 and 4 tended to agree. For infants, raters 2, 3, and 4 tended to agree.

Discussion

Agreement can be measured by using several statistics. Percentage agreement can provide an overall agreement rate; however, this statistic does not take chance into account. The κ statistic is a measure of agreement that indicates the proportion of agreement beyond that expected by chance.[4] The proportion of expected agreement is based on the assumption that assessments are independent between raters. To reflect the degree of disagreement, the κ statistic can be weighted so that greater emphasis is placed on large differences between ratings compared with small differences.[5] Weighted κ penalizes disagreements in terms of their seriousness, whereas unweighted κ treats all disagreements equally.[28] A number of weighting methods are available, but quadratic weighting is common.

Prior to the current study, anecdotal information suggested that when multiple raters assessed the same rhesus macaque, the scores were likely to agree within 0.5 score unit. A previous study similarly reported a high level of agreement (κ = 0.84 to 0.91) and agreement to within a half-score between instructors who had experience in BCS of cattle.[15] The current study demonstrated high percentage agreement (83% to 84%) and an overall substantial strength of agreement as determined by the κ coefficient (κ = 0.57 to 0.83) for adult and subadult animals.

Table 1. Number, age, and sex of macaques in each age category

		No. of animals by sex		
Age category	Actual age range	Male	Female	Total no. in group
Infant (<1 y)	11 d to 10 mo	45	69	115[a]
Juvenile (1–4 y)	0 y 11mo to 3 y 3 mo	104	149	253
Subadult (4–7 y)	3 y 11 mo to 6 y 3 mo	35	81	116
Adult (7–17 y)	6 y 11 mo to 17 y 3 mo	18	114	132

[a]Sex was recorded as 'unknown' for 1 infant

Table 2. Distribution of scores by age group

Age	Score								
group	1	1.5	2	2.5	3.0	3.5	4.0	4.5	5.0
Infant	17	146	**171**	108	18	0	0	0	0
Juvenile	0	23	372	**423**	190	4	0	0	0
Subadult	0	8	26	132	**195**	79	21	3	0
Adult	1	5	8	59	140	**167**	92	47	9

Data are given as the number of times the score was assigned; the most frequently assigned score (mode) is given in boldface.

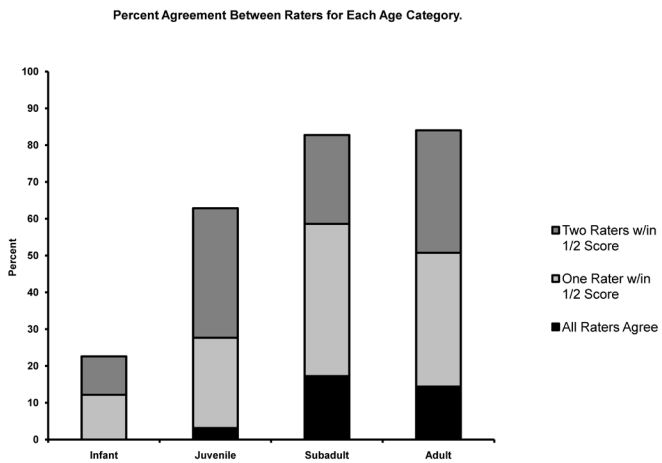

Figure 2. Percentage agreement between raters for each age category.

The frequency distribution of pairwise comparisons between raters demonstrated that most scores fell within one half of a score unit over all age groups.

Compared with that for adults and subadults, agreement between raters was less for juvenile animals and least for infants. Once data collection was completed, raters discussed their impressions of scoring different ages of macaques. In particular, raters indicated that scoring infants was particularly difficult, because they have so little muscle mass and fat reserves. Raters seemed to differ in their application of the score in regard to whether they scored an infant macaque 'as is' or as 'expected for the age.' In this sense, an infant with little muscle mass or fat deposits is essentially 'normal' and when scored for what is 'expected for age,' would be scored as 3.0, but when scored 'as is,' an infant might receive a score of only 2.0. Juvenile animals presented the same challenge, particularly because they are typically thin or lean and may experience periods of significant growth and change in body stature. Pairwise comparisons (Table 4) highlighted the effect of the scoring discrepancies. For example, for infants, one rater tended to score animals noticeably higher than did all 3 other raters, consistent with the postscoring discussion of scoring the infant animals as what is 'expected for age' rather than 'as is.' To further illustrate the

Table 3. κ values (with SE in parentheses) for each pairwise comparison between raters, representing interrater agreement

		Rater		
Age category	Rater	2	3	4
Infant	1	0.211 (0.036)	0.089 (0.016)	0.145 (0.037)
	2		0.389 (0.067)	0.190 (0.076)
	3			0.196 (0.051)
Juvenile	1	0.385 (0.061)	0.221 (0.027)	0.193 (0.023)
	2		0.256 (0.035)	0.169 (0.029)
	3			0.491 (0.062)
Subadult	1	0.764 (0.091)	0.715 (0.083)	0.580 (0.072)
	2		0.680 (0.083)	0.574 (0.073)
	3			0.749 (0.090)
Adult	1	0.800 (0.082)	0.793 (0.081)	0.759 (0.079)
	2		0.790 (0.084)	0.791 (0.083)
	3			0.829 (0.086)

*, $P < 0.05$ for all pairwise comparisons ($\kappa - 2SE > 0$)

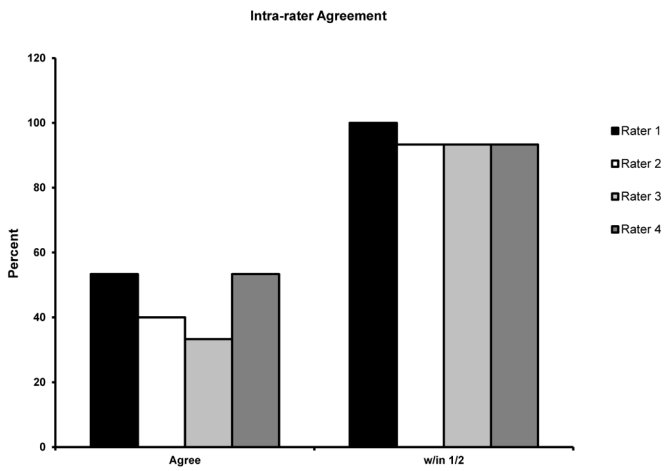

Figure 3. Intrarater agreement.

effect of 'expected' compared with 'as is' scoring, the overall percentage agreement to within one half a score unit for the infant group increased 7-fold when rater 1, who scored as expected, was removed from the calculation (data not shown). When this determination was repeated after individual exclusion of raters 2, 3, and 4, increases in agreement were only 3-, 4-, and 2-fold, respectively, and were similar among these 3 raters, who reported scoring macaques 'as is.' This example suggests the potential for increased agreement between raters scoring infants when the scoring system is applied as described.

Table 4. Frequency distribution of pairwise comparisons for raters over all scores by age group

		No. of times for which the difference between raters was						
Age group	Raters	–1.5	–1.0	–0.5	0	0.5	1.0	1.5
Adult								
	1 compared with 2	0	0	**16**	**69**	**41**	7	0
	1 compared with 3	0	0	**7**	**63**	**53**	7	2
	1 compared with 4	0	0	**10**	**48**	**64**	9	2
	2 compared with 3	0	1	**21**	**64**	**42**	4	0
	2 compared with 4	0	2	**13**	**66**	**48**	4	0
	3 compared with 4	0	3	**23**	**68**	**36**	2	0
Subadult	1 compared with 2	0	1	**13**	**77**	**23**	2	0
	1 compared with 3	0	0	**4**	**57**	**48**	7	0
	1 compared with 4	0	0	**2**	**39**	**61**	13	1
	2 compared with 3	0	0	**7**	**63**	**39**	7	0
	2 compared with 4	0	0	**3**	**44**	**59**	10	0
	3 compared with 4	0	0	**12**	**66**	**34**	4	0
Juvenile	1 compared with 2	0	1	**36**	**150**	**58**	8	1
	1 compared with 3	0	1	**0**	**48**	**163**	41	1
	1 compared with 4	0	0	**0**	**31**	**172**	49	1
	2 compared with 3	0	0	**6**	**66**	**147**	34	0
	2 compared with 4	0	0	**6**	**57**	**140**	50	0
	3 compared with 4	0	0	**30**	**170**	**53**	1	0
Infant	1 compared with 2	0	0	**1**	**14**	**51**	44	4
	1 compared with 3	0	0	**0**	**1**	**40**	69	5
	1 compared with 4	0	0	**3**	**16**	**72**	24	0
	2 compared with 3	0	0	**17**	**48**	**44**	6	0
	2 compared with 4	2	10	**35**	**45**	**22**	1	0
	3 compared with 4	0	6	**65**	**43**	**1**	0	0

Data are given as the difference of score for first rater minus that for the second rater listed.
Nos. of times that the agreement between raters was within 0.5 score unit are in boldface.

In light of these discrepancies, particularly those for the infant and juvenile macaques, we suggest that scoring each animal 'as is' would help to eliminate any differences between what individuals raters envision as 'ideal for age.' The resulting score might then be assessed in light of the macaque's age as well as other parameters, such as experimental manipulation, health status, and breeding status. In addition, if all raters scored animals relative to an age-appropriate ideal (BCS 3.0), individual scoring systems for each age group would need to used, likely leading to confusion. Even though particular age groups were associated with increased variability in scoring, looking at the scores assigned most frequently for each age group reveals that certain scores are more common for certain age groups. It then follows that the score assigned to a macaque would be interpreted in light of what would be a reasonable score for an animal of that age. Review of the frequency distribution of scores for each age group indicates that as age increased, the most frequently assessed body condition score increased. It is not unexpected that infants, on average, are thin (BCS 2.0), whereas adult macaques tend to be heavy (BCS 3.5).

Intrarater agreement to within one half of a score unit was high (93% to 100%), and κ values reflect substantial agreement. When presented with the same macaque within a short period of time, raters are likely to score the animal similarly. This outcome supports the usefulness of our scoring system. Good agreement between raters and within raters regarding BCS of nonhuman primates supports the validity of the scoring system. For the current study, the scoring system was applied most consistently for subadult and adult macaques. By scoring all age groups, including infants and juveniles, on an 'as is' basis, we anticipate that the agreement between raters over all age groups will improve. Raters who use our BCS system likely will be in agreement to within 0.5 score unit.

The goals for our BCS system are to provide a well-described scale that can be applied consistently and uniformly across raters, animals, and facilities. We envision that the score would be applied as part of the physical examination, to provide additional details with regard to the animal's body composition. The score would then be interpreted in light of the animal's status including age, nutrition, health, experimental use, and breeding. Various scores, particularly those that approach the extremes of the scale, might prompt nutritional, experimental, diagnostic, or therapeutic intervention. Our current study demonstrates that raters who score completely independently and are from different institutions show good agreement to within one half score. In an everyday setting in which multiple raters are scoring the same animal, assigning a consensus score likely would be the desired approach.

Acknowledgments

We gratefully acknowledge the Association for Primate Veterinarians for providing funding for this work. We thank the following people for

their assistance with this work: Beth Ford, Jamus MacGuire, and Shelly Lenz for assistance with data collection; James Koziol for biostatistical analysis; and the CNPRC research services staff, including Vanessa Bakula, Jennifer Vandevegte, Paul-Michael Sosa, and Ross Allen. This publication was made possible by grant 5P51 RR000169 from the National Center for Research Resources (NCRR) and by the NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

References

1. **Berman CM, Schwartz S.** 1988. A nonintrusive method for determining relative body fat in free-ranging monkeys. Am J Prim **14:**53–64.
2. **Burkholder WJ.** 2000. Use of body condition scores in clinical assessment of the provision of optimal nutrition. J Am Vet Med Assoc **217:**650–654.
3. **Clingerman KJ, Summers L.** 2005. Development of a body condition scoring system for nonhuman primates using *Macaca mulatta* as a model. Lab Anim (NY) **34:**31–36.
4. **Cohen J.** 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas **20:**37–46.
5. **Cohen J.** 1968. Weighted κ: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull **70:**213–220.
6. **Donoghue S, Khoo L, Glickman LT, Kronfeld DS.** 1991. Body condition and diet of relatively healthy older dogs. J Nutr **121:**S58–S59.
7. **Dorsten CM, Cooper DM.** 2004. Use of body condition scoring to manage body weight in dogs. Contemp Top Lab Anim Sci **43:**34–37.
8. **Foltz CJ, Ullman-Cullere M.** 1999. Guidelines for assessing the health and condition of mice. Lab Anim **28:**28–32.
9. **German AJ, Holden SL, Bissot T, Morris PJ, Biourge V.** 2008. Changes in body composition during weight loss in obese client-owned cats: loss of lean tissue mass correlates with overall percentage of weight lost. J Feline Med Surg **10:**452–459.
10. **German AJ, Holden SL, Bissot T, Morris PJ, Biourge V.** 2009. Use of starting condition score to estimate changes in body weight and composition during weight loss in obese dogs. Res Vet Sci **87:**249–254.
11. **Hickman DL, Swan M.** 2010. Use of a body condition score technique to assess health status in a rat model of polycystic kidney disease. J Am Assoc Lab Anim Sci **49:**155–159.
12. **Hoedemaker M, Prange D, Gundelach Y.** 2009. Body condition change ante- and postpartum: health and reproductive performance in German Holstein cows. Reprod Domest Anim **44:**167–173.
13. **Institute for Laboratory Animal Research.** 1996. Guide for the care and use of laboratory animals. Washington (DC): National Academies Press.
14. **Kovacs MS, McKiernan S, Potter DM, Chilappagari S.** 2005. An epidemiological study of interdigital cysts in a research beagle colony. Contemp Top Lab Anim Sci **44:**17–21.
15. **Kristensen E, Duehom L, Vink D, Andersen JE, Jakobsen EB, Illum-Nielsen S, Petersen FA, Enevoldsen C.** 2006. Within- and across-person uniformity of body condition scoring in Danish Holstein cattle. J Dairy Sci **89:**3721–3728.
16. **Kronfeld DS, Donoghue S, Glickman LT.** 1991. Body condition and energy intakes of dogs in a referral teaching hospital. J Nutr **121:**S157–S158.
17. **Kronfeld DS, Donoghue S, Glickman LT.** 1994. Body condition of cats. J Nutr **124:**2683S–2684S.
18. **Laflamme D.** 1997. Development and validation of a body condition score system for cats: a clinical tool. Feline Pract **25:**13–18.
19. **Laflamme D.** 1997. Development and validation of a body condition score system for dogs. Canine Pract **22:**10–15.
20. **Landis JR, Koch GG.** 1977. The measurement of observer agreement for categorical data. Biometrics **33:**159–174.
21. **McGregor BA, Butler KL.** 2008. Relationship of body condition score, live weight, stocking rate, and grazing system to the mortality of Angora goats from hypothermia and their use in the assessment of welfare risks. Aust Vet J **86:**12–17.
22. **Parker R.** 1996. Using body condition scoring in dairy herd management. Guelph (Canada): Ontario Ministry of Food and Agriculture Factsheet.
23. **Paster EV, Villines KA, Hickman DL.** 2009. Endpoints for mouse abdominal tumor models: refinement of current criteria. Comp Med **59:**234–241.
24. **Roche JF.** 2006. The effect of nutritional management of the dairy cow on reproductive efficiency. Anim Reprod Sci **96:**282–296.
25. **Rodenburg J.** 1996. Body condition scoring of dairy cattle. Guelph (Canada): Ontario Ministry of Food and Agriculture Factsheet.
26. **Russel A.** 1984. Body condition scoring of sheep. In Pract **6:**91–93.
27. **Scarlett JM, Donoghue S.** 1998. Associations between body condition and disease in cats. J Am Vet Med Assoc **212:**1725–1731.
28. **Sim J, Wright CC.** 2005. The κ statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther **85:**257–268.
29. **Slupe JL, Freeman LM, Rush JE.** 2008. Association of body weight and body condition with survival in dogs with heart failure. J Vet Intern Med **22:**561–565.
30. **Wright B, Rietveld G, Lawlis P.** 1998. Body condition scoring of horses. Guelph (Canada): Ontario Ministry of Food and Agriculture Factsheet.